

MyIMG 4.0

1 Synopsis

IMG supports MyIMG and missing gene annotations for users to add or update existing gene information and to facilitate community annotations. This document describes MyIMG annotations, My Missing Genes annotations and the new IMG Group extension to support better group and community genomic annotations. Important new features include:

- Users can now form their own IMG Groups to share annotations.
- Users can participate in multiple IMG Groups.
- Users can selectively share annotations with different groups.

IMG Group, MyIMG and Missing Gene Annotations requires a login and password which can be requested by filling the **Request Account** form at: <http://img.jgi.doe.gov/request>.

2 IMG Groups

In order to view, add, or modify IMG groups, a user should first log into IMG ER or MER and to find **MyIMG Home** menu item under the **MyIMG** tab. If the user already belongs to an IMG group, the information will be displayed. For most users who do not belong to any IMG group, the display will show "No IMG Groups" (Figure 2.1):

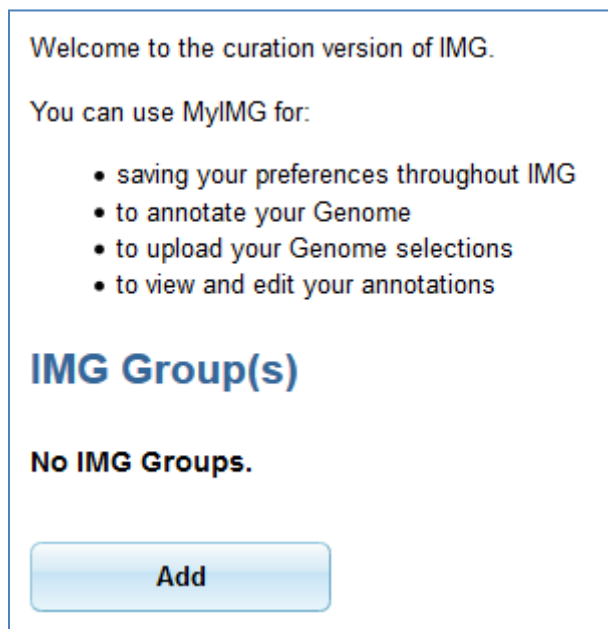


Figure 2.1: IMG Groups

In order to create a new IMG group, click the "Add" button, and then fill the information in the following page (Figure 2.2):

Add A New IMG Group

Group Information

Group Name	My Annotation Group
Owner	Amy Test Data (amy_test)
Owner Email	IMACHen@lbl.gov
Description	my group

[Add Group](#)

Figure 2.2: Create New IMG Group

After a group is created, the owner can update the group to add members by selecting the group and click the "Add" button:

IMG Group(s)

Filter column: Filter [Apply](#) [?](#)

[Export](#) Page 1 of 1 << first < prev 1 next > last >> All

[Column Selector](#) [Select Page](#) [Deselect Page](#)

Select	Group ID	Group Name	Owner Name	Add Date	Description	Member Count	My Role
<input type="radio"/>	27	My Annotation Group	Amy Test Data	2015-03-16	my group	1	owner

[Export](#) Page 1 of 1 << first < prev 1 next > last >> All

[Add](#) [Update](#) [Delete](#)

Figure 2.3: IMG Group with Group Information

The owner of a group can perform the following update functions:

- update group description;
- add new members to the group;
- change certain member roles to "co-owner" to help managing the group;
- remove members that are no longer belonging to this group.

Group Information

Group ID	27
Group Name	My Annotation Group
Owner	Amy Test Data (amy_test)
Owner Email	IMACHen@lbl.gov
Add Date	2015-03-19
Description	my group

Save Update in Database

(Click the Save button after you make changes to group name and/or description; otherwise the change won't be saved in database)

Members

News

Members

No members in this group

Change selected to: co-owner

Update Member Role

Delete Member(s)

Add New Members

IMG user names or JGI SSO user names (separated by ,):

Add Member(s)

Figure 2.4: Update IMG Group Information

If a user is a member (but not owner) of any IMG groups, there will be a "Withdraw" button in Figure 2.3 for the user to withdraw from the group. Owners cannot withdraw from their groups, but they can select to delete a group.

Group members can also share news. In order to view or post group news, click the "News" tab (Figure 2.5). Group news are considered private to group members only. However, certain IMG Datamarts can announce public group news to general users to the datamarts.

Members News

News

No Group News.

Post News

Title	<input type="text"/>
Description	<input type="text"/>
Is Public?	No ▾

Post News

Figure 2.5: IMG Group News

Group members can also grant genome access permissions to other members in the same group. However, please note that:

- only PIs and submitters of the genomes have the privilege to grant access of their genomes;
- there is no web function to revoke access permissions. (Contact us if you need to revoke access permissions.)

Share Genomes with Group Members

Note: This option is for new sharings only. It does not revoke previously granted access permissions. Contact us if you wish to revoke any access permission.

Please enter IMG OIDs (separated by ,):

Grant Access

Figure 2.6: Share Genomes with Group Members

3 MyIMG Curation of Functional Annotations

In order to view, add, or modify MyIMG annotations, a user should find the **Annotations** menu item under the **MyIMG** tab (Figure 3.1):

IMG User Annotations

You can view annotations, or upload annotations from flat files.

View Annotations

☒ View all annotations
☐ View annotations by genomes
☐ View annotations for all genes in gene cart

[View My Annotations](#)
[View Group Annotations](#)

Upload Annotations from File

Upload your own annotations from a tab-delimited file.

[Upload Annotations](#)

View My Missing Gene Annotations

[View My Missing Genes](#)
[View Group Missing Genes](#)

Figure 3.1: IMG User Annotations

There are 3 sections on this page:

- **View Annotations:** view, add, or update user's MyIMG annotations. There is also a "View Group Annotations" button if the user belongs to an IMG group.
- **Upload Annotations from File:** update MyIMG annotations from a tab-delimited file.
- **View My Missing Gene Annotations:** view, add, or update user's missing gene annotations. There is also a "View Group Missing Genes" button if the user belongs to an IMG group.

The functional annotation for individual genes can be curated using the **MyIMG Annotations** features of **MyIMG**. In addition to curation of functional annotations, **MyIMG** provides support for uploading user genome selections that have been saved earlier from **Genome Browser** or **Genome Statistics** and set system wide user preferences.

MyIMG Annotations provides support for: (1) **editing** the product name and several associated fields for one or more related genes that have been previously selected and saved in the **Gene Cart**; (2) finding genes associated with **enzymes** missing in specific genomes; (3) **reviewing** the functional annotations for individual genes or group of genes; (4) associating genes with additional **IMG terms**; (5) marking genes **obsolete**; (6) **exporting to/ uploading** from a tab-delimited file functional annotations for genes identified by their IMG identifier (OID).

3.1 Editing Product Names

Genes that may require **product name review** and **curation** can be identified using various analysis tools. For example, genes without a product name but with evidence of potential functional annotation or with product name but without any evidence of functional annotation are candidates for product name review and curation.

MyIMG provides support for editing the product name and associated information for one or several related genes. The following annotation fields can be manually edited:

- Product Name
- Protein Description
- EC Number
- PUBMED ID
- Inference
- Is Pseudo Gene?
- Notes
- Gene Symbol

First, select and save the target genes for product name curation in the **Gene Cart**. Next, use the **Annotate Selected Genes** link in the **MyIMG Annotation** section of the **Gene Cart** to access the tool that allows editing the fields listed above.

Example 3.1. Consider for review gene PF1186 (IMG identifier 638173757) of genome *Pyrococcus furiosus* whose details are shown in Figure 3.1(i). This gene is associated with product name NADH oxidase, as shown in Figure 3.1(i), and as recorded in GenBank and RefSeq which can be viewed by following the appropriate External Links. The list of top homologs for the gene under review can be displayed via the Homologs section of its **Gene Details**, as shown in Figure 3.1(ii).

Gene Details (i)

Gene Information	
Gene Object ID	638173757
Gene Symbol	PF1186
Locus Tag	PF1186
Product Name	NADH oxidase
Description	Function Code: 4.5 Cellular Process
Genome	Pyrococcus furiosus DSM 3638
DNA Coordinates	1131551..1132879 (-)(1329bp)
Scaffold Source	Pyrococcus furiosus DSM 3638: A
IMG ORF Type	
GC Content	0.44
Accession	AAL81310
External Links	GI:18893271 ; PUMA/GI: 8893271
Fusion	No

Top IMG Homolog Hits (ii)

☒ Add query gene [638173757](#)

Select	Homolog	I	Product Name	Percent Identity	Alignment On Query Gene	Alignment On Subject Gene	Length
<input checked="" type="checkbox"/>	638207469	-	NADPH:sulfur oxidoreductase (MyIMG: gbp)	88.24			442aa
<input checked="" type="checkbox"/>	638200851	-	NADPH:sulfur oxidoreductase (MyIMG: gbp)	85.68			446aa
<input checked="" type="checkbox"/>	638186008	-	NADPH:sulfur oxidoreductase (MyIMG: gbp)	85.45			445aa

Gene Cart (iii)

Remove Selected Select All Clear All

Selection	Gene Object ID	Locus Tag	Product Name
<input checked="" type="checkbox"/>	638173757	PF1186	
<input checked="" type="checkbox"/>	638207469	TK1299	
<input checked="" type="checkbox"/>	638186008	PH0572	
<input checked="" type="checkbox"/>	638200851	PAB0936	

Enter MyIMG Annotation

You may enter, update or delete your product name, function, EC number, P

MyIMG Annotation for Selected Genes (iv)

Select	Gene Object ID	Locus Tag	Original Product Name	Annotated Product Name	AA Seq. Length	Genome
<input checked="" type="checkbox"/>	638173757	PF1186	NADH oxidase	NADPH:sulfur oxidoreductase	442aa	<i>Pyrococcus furiosus</i> DSM 3638
<input checked="" type="checkbox"/>	638186008	PH0572	445aa long hypothetical NADH oxidase	NADPH:sulfur oxidoreductase	445aa	<i>Pyrococcus horikoshii</i> OT3
<input checked="" type="checkbox"/>	638200851	PAB0936	NADH oxidase (noxA-1)	NADPH:sulfur oxidoreductase	446aa	<i>Pyrococcus abyssi</i> GE5
<input checked="" type="checkbox"/>	638207469	TK1299	NADH oxidase	NADPH:sulfur oxidoreductase	442aa	<i>Thermococcus kodakarensis</i> KOD1

MyIMG Annotation

Product Name:

Prot Description:

EC Number:

PUBMED ID:

Inference:

Notes:

Is Pseudo Gene?

Update Annotation Delete Annotation Reset Cancel

Figure 3.1. Review and curation of the product name for a gene of *Pyrococcus furiosus* using **MyIMG** Annotation.

Based on a recent study¹, it has been determined that the function for this gene is NADPH:sulfur oxidoreductase, and an expert review of the best homologs of this gene indicated that this product name also may be confidently applied to the top three homologs. The gene under review and these top homologs are added to the **Gene Cart**, as shown in Figure 3.1(iii).

Next, the product name where is changed to NADPH:sulfur oxidoreductase using the **MyIMG Annotation** tool accessed from **Gene Cart**, as shown in Figure 3.1(iv). Other annotations (e.g., EC number) can be also modified. User annotations are stored in IMG and can be reviewed at any time using **MyIMG** viewing options, as shown in Figure 3.2.

¹ Schut, G.J., Bridger, S.L., Adams, M.W. (2007) Insights into the metabolism of elemental sulfur by the hyperthermophilic archaeon *pyrococcus furiosus*: characterization of a coenzyme a-dependent NAD(P)H Sulfur Oxidoreductase. *Journal of Bacteriology*.

3.2 Finding Missing Enzymes

The metabolic capacity of a genome is defined by its association with pathways via gene products that function as enzymes serving as catalysts for reactions in these pathways. A genome-pathway association may be partial, with “missing” associations between enzymes for reactions on the pathway with genes on the genome. We call such associations **missing enzymes**.

3.2.1 Missing Enzymes for Specific Genomes and Genes

MyIMG provides support for revising **missing enzymes** for specific genomes and genes. For each genome, the **Genome Statistics** section of its **Organism Details** page contains a count of “Genes without enzymes, but with PRIAM hits” which leads to a list of genes that could be associated with enzymes predicted by PRIAM. These predicted enzymes can be examined for accuracy and then associated with genes using **MyIMG Annotation** tools.

Example 3.2. Use the **Genome Browser** for selecting *Thermoplasma volcanium* GSS1 (*T. volcanium*) genome and examine the **Genome Statistics** section of its **Organism Details**, as shown in Figure 3.2(i). Follow the “Genes w/o enzymes but with candidate KO based enzymes” link to the list of genes that have candidate KO based enzymes, as shown in Figure 3.2(ii).

The screenshot displays the MyIMG Genome Browser interface for the *Thermoplasma volcanium* GSS1 genome. The interface is divided into several sections:

- New Annotations (iii):** A section for adding or updating annotations, including buttons for "Change MyIMG Annotations", "Add/Update IMG Term(s)", "Add To Gene Cart", "Export Annotations", "Select All", and "Clear All".
- Genome Statistics (i):** A table showing various genomic metrics for the *Thermoplasma volcanium* GSS1 genome. The table includes columns for "Number" and "% of Total".
- Genes w/o enzymes but with candidate KO based enzymes (ii):** A table listing genes that have candidate KO-based enzymes but no associated enzymes. The table includes columns for "Gene ID", "Product Name", "KO ID", "KO Definition", "Percent Identity", "Alignment On Gene", "E-value", and "Bit Score".

Red arrows highlight the "Genes w/o enzymes but with candidate KO based enzymes" link in the top left and the "Genes w/o enzymes but with candidate KO based enzymes" table in the bottom right.

Category	Number	% of Total
DNA, total number of bases	1584804	100.00%
DNA coding number of bases	1399584	88.31%
DNA G+C number of bases	632630	39.92% ¹
DNA scaffolds	1	100.00%
CRISPR Count	2	
genes total number	1615	100.00%
Protein coding genes	1561	96.66%
Pseudo Genes	64	3.96% ²
rRNA genes	54	3.34%
rRNA genes	3	0.19%
5S rRNA	1	0.06%
16S rRNA	1	0.06%
23S rRNA	1	0.06%
tRNA genes	46	2.85%
Other RNA genes	5	0.31%
Protein coding genes with function prediction	1092	67.62%
without function prediction	469	29.04%
Protein coding genes connected to SwissProt Protein Product	95	5.88%
not connected to SwissProt Protein Product	1466	90.77%
Protein coding genes connected to SEED	931	57.65%
not connected to SEED	630	39.01%
Protein coding genes with enzymes	462	28.61%
w/o enzymes but with candidate KO based enzymes	15	0.93%

Figure 3.2. Examining Missing Enzymes for a Specific Genome.

Select the genes you want to associate with the predicted enzymes and then use **Update MyIMG Annotation**, as shown in Figure 3.2(ii). You can either **Add** the predicted enzyme to or **Replace** an existing enzyme in your MyIMG annotation. The new gene-enzyme associations are listed for review, as shown in Figure 3.2(iii).

Instead of associating lists of genes with the candidate KO based enzymes, one can examine all enzyme predictions for individual genes via **Find Candidate Enzymes with PRIAM** available in the **Gene Information** section of the gene's **Gene Detail** page, as illustrated in Figure 3.3(i).

Example 3.3. In the list of genes of *Thermoplasma volcanium* GSS1 (*T. volcanium*) genome that have candidate KO based enzymes (see Figure 3.2 (ii)) go to the **Gene Detail** page for the gene with IMG identifier 638191966. Use **Find Candidate Enzymes** immediately before the **Find Candidate Product Name** section, as shown in Figure 3.3(i), to get to the list predicted genes on the **Candidate Enzymes Using Kegg Ontology (KO)** page, as shown in Figure 3.3(ii). Examine the enzymes in the list and select the best predicted enzyme(s). Use **Add to MyIMG Annotation**, as shown in Figure 2.3(ii), and then either select **Add** or **Replace**, as shown in Figure 3.3(iii), to update the MyIMG enzyme annotation for this gene.

Gene Information (i)

Gene Information	
Gene ID	638191966
Gene Symbol	TVG1420982
Locus Tag	TVG1420982
Product Name	hypothetical protein
SwissProt Protein Product	
SEED	
IMG Term	chorismate mutase (EC 5.4.99.5) (Iain Anderson 2005-04-28)
Genome	Thermoplasma volcanium GSS1
DNA Coordinates	1420446..1420982 (+)(537bp)
Scaffold Source	Thermoplasma volcanium GSS1 DNA: BA000000

[Add To Gene Cart](#) [Find Candidate Enzymes](#) [Show All](#)

Candidate Enzymes Using Kegg Ontology (KO) (ii)

Gene (638191966): hypothetical protein

Filter column: Candidate Enzyme Filter text Apply

Export Page 1 of 1 << first < prev 1 next > last >> All

Select	Candidate Enzyme	Enzyme Name	KO ID	KO Definition	Enzymes associated with this KO
<input checked="" type="checkbox"/>	EC:1.3.1.12	Prephenate dehydrogenase.	KO:K14187	chorismate mutase / prephenate dehydrogenase [EC:5.4.99.5 1.3.1.12]	EC:1.3.1.12, EC:5.4.99.5
<input checked="" type="checkbox"/>	EC:5.4.99.5	Chorismate mutase.	KO:K14187	chorismate mutase / prephenate dehydrogenase [EC:5.4.99.5 1.3.1.12]	EC:1.3.1.12, EC:5.4.99.5

Export Page 1 of 1 << first < prev 1 next > last >> All

[Add to MyIMG Annotation](#) [Select All](#) [Clear All](#)

Add Enzyme(s) to Selected Gene in MyIMG Annotation (iii)

Gene (638191966): hypothetical protein

Select	EC Number	Enzyme Name
<input checked="" type="checkbox"/>	EC:1.3.1.12	Prephenate dehydrogenase.
<input checked="" type="checkbox"/>	EC:5.4.99.5	Chorismate mutase.

Add or replace MyIMG gene-enzyme annotation:
☒ Add ☐ Replace

hint: Click [Update MyIMG Annotation](#) to save your change(s) to the database. Only gene-enzyme associations will be added to MyIMG annotation.

[Update MyIMG Annotation](#) [Reset](#)

Figure 3.3. Examining Missing Enzymes for a Specific Gene.

3.2.2 Missing Enzymes within a KEGG Pathway or Function Profile

MyIMG provides support for **examining missing enzymes** for a specific genome, **G**, using either a **KEGG Pathway Map** for **G** or a **Functional Profile** involving **G** and enzymes of interest, as illustrated in Figure 3.4.

Example 3.4. Use the **Genome Browser** for selecting *Thermoplasma volcanium* GSS1 (*T. volcanium*) and *Thermoplasma acidophilum* DSM 1728 (*T. acidophilum*) genomes. Save these selections. Select **Find Functions** in the Main Menu and then select the **KEGG** browser and **Pathways via EC Numbers** submenu item, as illustrated in Figure 3.4(i). Selecting the *Fatty acid metabolism* pathway under **Lipid Metabolism** will lead to the **KEGG Pathway Details** page for this pathway, as shown in Figure 3.4(ii). You can either:

- (i) Select the **View Map for Selected Genomes** tab, and then select *T. volcanium* and "Find missing enzymes" option, and click **View Map** at the bottom of the **KEGG Pathway Details** to display the KEGG map for the *Fatty Acid metabolism* pathway, as shown in Figure 3.4(iii), or
- (ii) Select enzymes of interest from the list of enzymes (e.g., EC:1.3.3.6, EC:1.3.99.3, and EC:6.2.1.3) provided by the **KEGG Pathway Details**, save them in **Function Cart**, and compute a **Function Profile** for these enzymes across *T. volcanium* and *T. acidophilum* which will result in the profile shown in Figure 3.4(iv).

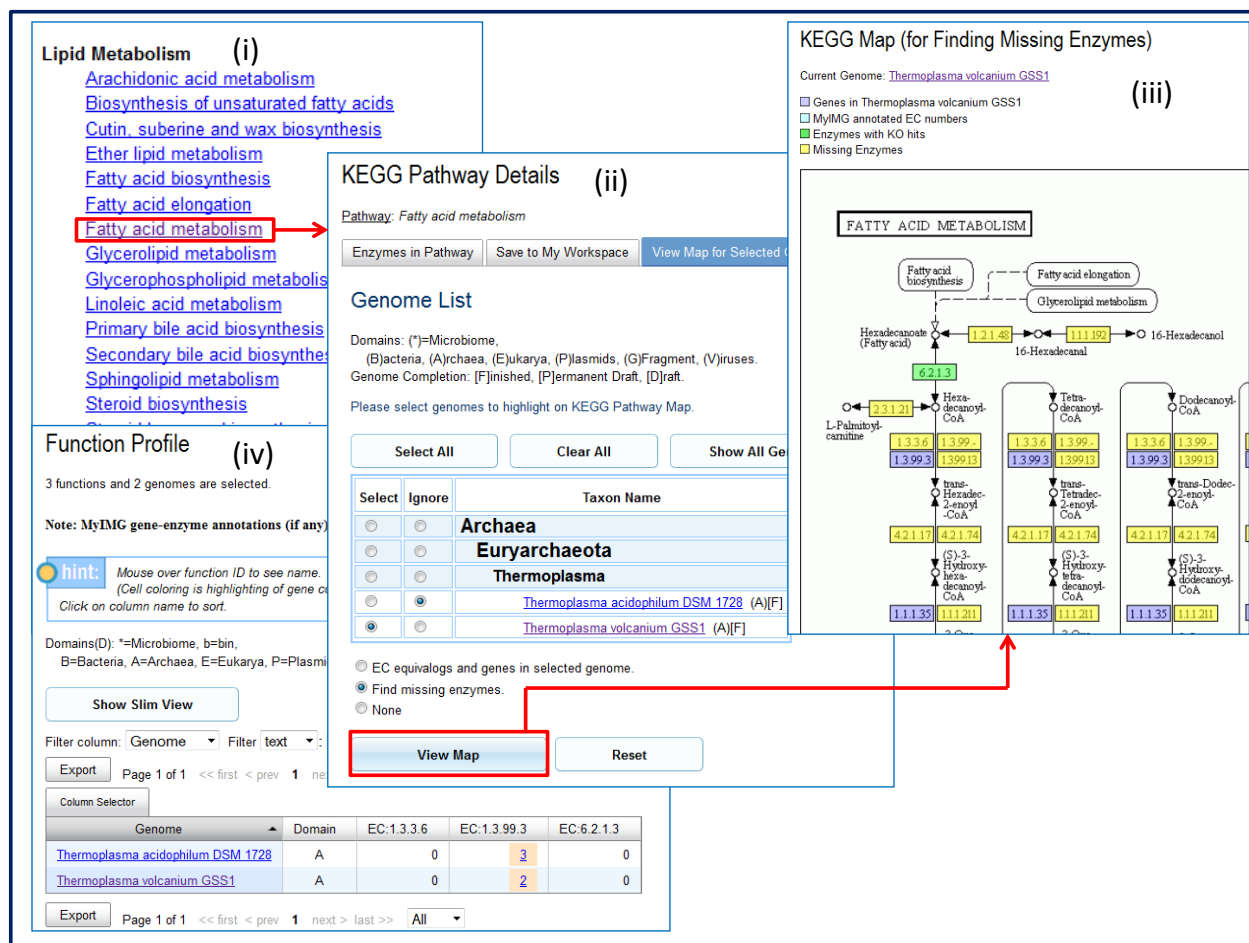


Figure 3.4. Examining Missing Enzymes with KEGG Pathway Map and Function Profile.

On the **KEGG Map**, enzymes that are associated with a *T. volcanium* gene are colored **blue**, enzymes that are associated with any *T. volcanium* MyIMG gene annotations are colored **cyan**, while so called “missing” enzyme are colored either **green**, for enzymes that have candidate enzymes with KO hits to a *T. volcanium* gene, or **yellow**, for enzymes without any KO hits to *T. volcanium* genes. Clicking on a missing enzyme such as EC:6.2.1.3, as illustrated in Figure 3.5(i), will lead to a **Find Candidate Genes for Missing Function** page, as shown in Figure 3.5(iii). Note that selection of a (green colored) missing enzyme that has a KO hit enhances the chances of finding for it good candidate genes.

In the **Function Profile** result, enzymes that are associated with *T. volcanium* or *T. acidophilum* genes are identified by positive integer numbers which represent the count of genes associated with the enzymes, while so called “missing” enzyme are identified by a “0”.

For missing enzymes of interest, **MyIMG** provides support for finding **candidate genes** that could be associated with these enzymes, as illustrated in Figures 3.3 and 3.4.

Example 3.5. Consider “missing” enzyme **EC:6.2.1.3** identified as discussed in Example 3.4 above, and leading to **Find Candidate Genes for Missing Function** page, as shown in Figure 3.5(iii).

KEGG Map (for Finding Missing Enzymes)

Current Genome: [Thermoplasma volcanium GSS1](#)

☐ Genes in Thermoplasma volcanium GSS1
☐ MyIMG annotated EC numbers
☒ Enzymes with KO hits
☐ Missing Enzymes

FATTY ACID METABOLISM

Fatty acid biosynthesis
 Fatty acid elongation
 Glycerolipid metabolism
 Hexadecanoate (Fatty acid)
 16-Hexadecanal
 16-Hexadecanoate
 Hexadecanoyl-CoA
 Tetradecanoyl-CoA
 Dodecanoyl-CoA
 L-Palmitoyl-CoA

EC:6.2.1.3

Find Candidate Genes for Missing Function

[Thermoplasma volcanium GSS1](#)

Archaea; Euryarchaeota; Thermoplasma; Thermoplasmales; Thermoplasmales; Thermoplasma; volcanium

Function: (EC:6.2.1.3) Long-chain-fatty-acid--CoA ligase.

☒ Using Homologs
☐ Using KO
☐ Using Both

Go **Reset**

Using Homologs

This tool allows you to find genes associated with enzymes through homologs in other genomes. Homologs from the query genome *Thermoplasma volcanium GSS1* has homologs in other genomes associated with EC:6.2.1.3. These homologs have a reciprocal hits in the query genome, which are listed as candidates for gene-enzyme association.

Database Search Options:

☐ Currently selected genomes (fast)
☐ Whole database (slow)
☐ Archaea
☒ Euryarchaeota

Candidate Genes for Missing Function

Genome: [Thermoplasma volcanium GSS1](#)

Function: (EC:6.2.1.3) Long-chain-fatty-acid--CoA ligase.

Domains(D): ~Microbiome, B-Bacteria, AnArchaea, E-Eukarya, P-Plasmids, G-GFragment, V-Viruses, Genome Completions(C), F-Finished, P-Permanent Draft, D-Draft.

Filter column: Candidate Gene Filter text: Apply

Export Page 1 of 1 << first < prev 1 next > last >> All

Select	Candidate Gene	Candidate Gene Product	Enzyme for Candidate Gene	Homolog Gene	Homolog Gene Product
<input type="checkbox"/>	538190619	hypothetical protein	EC:6.2.1.1	2509440692	Long-chain-acid-CoA ligase
<input type="checkbox"/>	538191204	medium-chain acyl-CoA ligase		2502086732	Long-chain-acid-CoA ligase
<input type="checkbox"/>	538191754	acetyl-CoA synthetase	EC:6.2.1.1	2502086732	Long-chain-fatty-acid-CoA ligase

Figure 3.5. Finding Candidate Genes for **Missing Enzymes** on a **KEGG Pathway Map** using **Homologs**.

You can find candidate genes of *T. volcanium* that could be associated with this enzyme as follows:

- (i) **Search** for *T. volcanium* genes that have **homologs/orthologs** associated with enzyme EC:6.2.1.3, as illustrated in Figure 3.5(iii). You can search across all the genomes available in the system, across a subset of genomes within a certain domain/phyla/class, or only across the selected genomes (i.e. *T. acidophilum*). If you have started from a **Function Profile**, you can also restrict the search to the genomes involved in the profile. You can change the default values set for percent identity and e-value cutoffs and the number of retrieved homologs. The result of this search lists *T. volcanium* candidate genes, as illustrated in Figure 3.5(iv). You can select a candidate gene and associate it with the “missing” enzyme using the **MyIMG Annotation** tool.
- (ii) Employ **KO** for finding *T. volcanium* genes that could be associated with the “missing” enzyme, as illustrated in Figure 3.6(ii). You can change the default values set for percent identity, e-value, and percent alignment cutoffs.

The result of this search lists *T. volcanium* candidate genes, as illustrated in Figure 3.6(iii). You can select a candidate gene and associate it with the “missing” enzyme using the **MyIMG Annotation** tool.

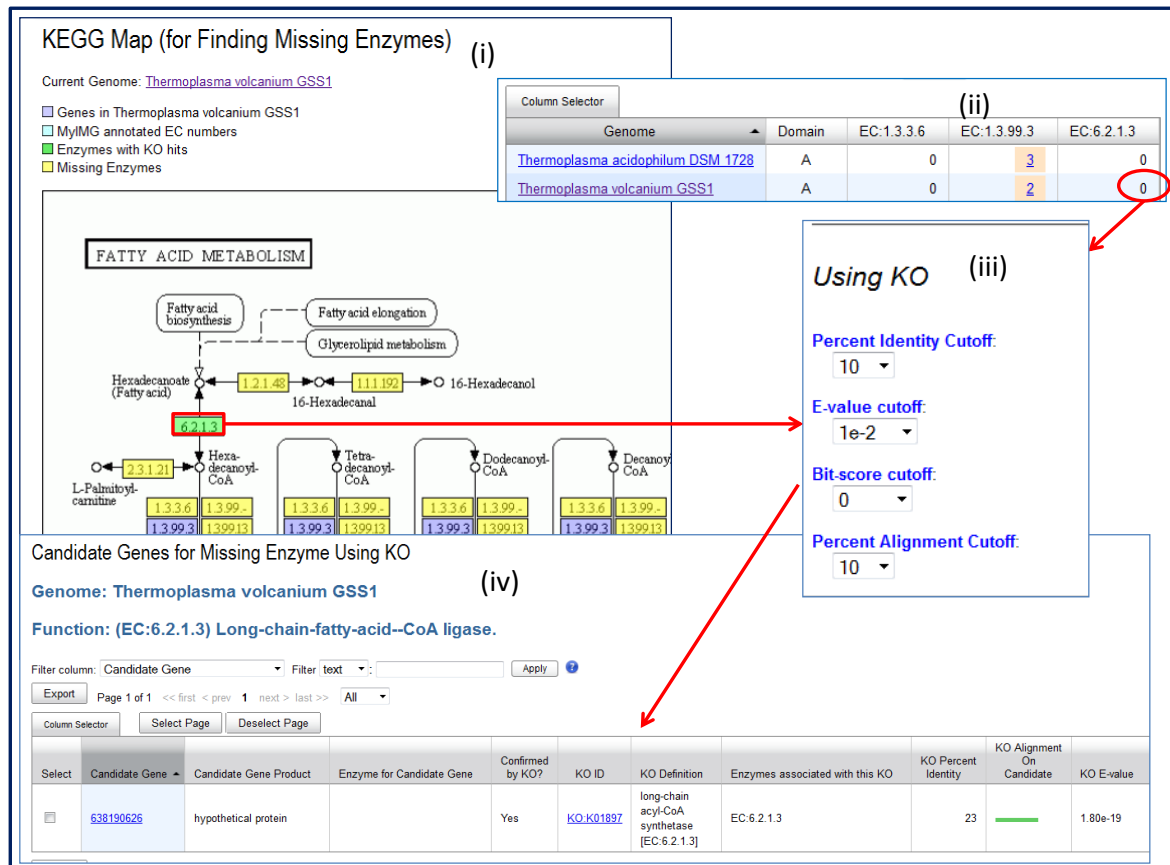


Figure 3.6. Finding Candidate Genes for Missing Enzymes on a KEGG Pathway Map using KO.

(iii) Employ a combination of **homolog/ortholog** based search and **KO**, which will list the results together.

From the list of candidate genes (e.g., see Figure 3.6(iv)), you can select a candidate gene and go to the **Add Enzyme to Candidate Gene(s) in MyIMG Annotation** tool shown in Figure 3.7(i). You can either add or replace the “missing” enzyme for the selected candidate gene. After the addition of MyIMG annotation, KEGG Map display will show enzyme EC:6.2.1.3 being color **cyan** instead of the original **green** (Figure 3.7 (ii)). Gene Detail of gene 638190626 (TVG0135483) page will display the new MyIMG annotation (Figure 3.7(iii)).

Add Enzyme to Candidate Gene(s) in MyIMG Annotation (i)

Select	Gene ID	Gene Display Name	Genome	Old MyIMG Enzyme(s)	New MyIMG Enzyme	Add/Replace
<input checked="" type="checkbox"/>	638190626	hypothetical protein	Thermoplasma volcanium GSS1		EC:6.2.1.3	<input checked="" type="radio"/> Add <input type="radio"/> Replace

hint:

Click 'Update MyIMG Annotation' to save your change(s) to the database. Only selected gene-enzyme associations will be added to MyIMG

Update MyIMG Annotation

Reset

KEGG Map (for Finding Missing Enzymes) (ii)

Current Genome: [Thermoplasma volcanium GSS1](#)

- Genes in Thermoplasma volcanium GSS1
- MyIMG annotated EC numbers
- Enzymes with KO hits
- Missing Enzymes

FATTY ACID METABOLISM

Gene Detail (iii)

MyIMG Annotation	
Product Name	hypothetical protein
EC Number	EC:6.2.1.3
Is Public?	No

Figure 3.7. Associating a Candidate Gene with an Enzyme using **MyIMG Annotation**.

3.3 Reviewing MyIMG Annotations

MyIMG annotations for all the genes that you have curated can be reviewed using the **View My Annotations** section of the **IMG User Annotations** page. This page can be accessed using the **Annotations** sub-menu of the **MyIMG** main menu tab, as illustrated in Figure 3.8(i). Three review alternatives are available:

- (i) All the genes can be displayed in a tabular format, where each row consists of the annotations for an individual gene, as illustrated in Figure 3.8(ii).
- (ii) The genes are first displayed grouped per genomes, as illustrated in Figure 3.8(iii). You can select the list of genes for a specific genome to review their annotations.
- (iii) Display is limited to only genes in gene cart.

Panel (i): IMG User Annotations

You can view annotations, or upload annotations from

View Annotations

- ☐ View all annotations
- ☒ View annotations by genomes
- ☐ View annotations for all genes in gene cart

View My Annotations **View Gro**

Upload Annotations from File

Upload your own annotations from

Upload Annotations

Panel (ii): My Annotations

Change MyIMG Annotations Add/Update IMG Term(s)

Add To Gene Cart Export Annotations Select All Clear All

Click on column name to sort.

Filter column: Genome Filter text: Apply

Export Page 1 of 1 << first < prev 1 next > last >> All

Select	Gene ID	Genome	Original Product Name	Annotated Product Name	Annotated Prot Desc	Annotated EC Number
<input type="checkbox"/>	637682586	Dechloromonas aromatica RCB	ATP synthase subunit B	ATP synthase subunit beta Q477Z1		EC:3.6.3.14
<input type="checkbox"/>	637258320	Fusobacterium nucleatum ATCC 25586	2-amino-4-hydroxy-6-hydroxymethylidihydropteridine pyrophosphokinase	2-amino-4-hydroxy-6-hydroxymethylidihydropteridine pyrophosphokinase		
<input type="checkbox"/>	638160826	Halobacterium sp. NRC-1	ribose phosphate pyrophosphokinase	ribose-phosphate pyrophosphokinase		
<input type="checkbox"/>	638202835	Methanocaldococcus jannaschii DSM 2661	hypothetical protein	shikimate kinase		

View my annotations on genome(s). Select one or more genomes and click the "View Annotations on Genome(s)"

Panel (iii): My Annotations

View My Annotations on Genome(s) Select All Clear All

Select	Taxon ID	Genome Name	Genes
<input type="checkbox"/>	637000088	Dechloromonas aromatica RCB	1
<input type="checkbox"/>	637000117	Fusobacterium nucleatum subsp. nucleatum ATCC 25586	1
<input type="checkbox"/>	638154504	Halobacterium sp. NRC-1	1
<input type="checkbox"/>	638154505	Methanocaldococcus jannaschii DSM 2661	1
<input type="checkbox"/>	638154508	Methanosarcina acetivorans C2A	1
<input type="checkbox"/>	638154509	Methanosarcina mazei Go1	1

Figure 3.8. View MyIMG Annotations.

The page displaying **MyIMG Annotations** for curated genes provides support for **uploading** annotations to a tab-delimited file using **Upload Annotations** (see Figure 3.8(i)). The file has the following column headers:

- Gene ID or OID (**required**): gene object ID
- Annotated Product Name (**required**): my annotated product name(s); separate multiple product names using ','

- Annotated Prot Desc (optional): my annotated prot description
- Annotated EC Number (optional): my annotated EC number(s); separate multiple EC numbers using space or ','
- Annotated PUBMED ID (optional): my annotated PUBMED ID(s); separate multiple PUBMED ID's using space or ','
- Inference (optional): my annotated inference
- Is Pseudo Gene? (optional): is pseudo gene? (Yes, No)
- Notes (optional): my annotated free text notes
- Annotated Gene Symbol (optional): my annotated gene symbol
- Remove Gene from Genome? (optional): remove gene from genome? (Yes, No)
- Is Public? (optional): is this annotation public? (Yes, No)

3.4 Group and Community Annotations

If a user belongs to an IMG group, then the user can view all MyIMG annotations by group members with the following restrictions:

- The user must have access permission to the genomes. All MyIMG annotations on private genomes will only be visible to other group members that have access permission to the private genomes. MyIMG annotations on public genomes are not restricted.
- Those MyIMG annotations must be either public or shared by the authors of the annotations. (An author can selectively share MyIMG annotations with different groups; e.g., sharing with Group 1, but not with Group 2. In this case, Group 1 members can view the MyIMG annotations, but Group 2 members cannot.)
- The user can only view, but not modify, MyIMG annotations by other group members.

To view MyIMG annotations by group members, click the "View Group Annotations" button in the IMG User Annotations page shown in Figure 2.1. Many IMG users have used this feature for group annotations with colleagues. For users who belong to multiple IMG groups, there will be a dropdown selection for users to switch groups.

There are two ways for a user to change the sharing options of his/her MyIMG annotations:

1. To share all MyIMG annotations of selected genome(s), a user should select "View annotations by genomes" and click the "View My Annotations" button (Figure 3.1). In the following **My Annotations by Genomes** page, the user selects a subset of genomes and then specify the sharing option in the "Update My Annotation Sharing in Selected Genome(s)" section.
2. To share individual MyIMG annotations, a user should select "View all annotations" and click the "View My Annotations" button (Figure 3.1). Select all the annotations one wishes to share, and then click the "Change MyIMG Annotations" button. In the following updating annotation page, there is a new section for the user to specify sharing option.

MyIMG annotations are private by default. However, IMG users can change any of their MyIMG annotations to public. Public MyIMG annotations are visible to all IMG users provided that users

have access permissions to the corresponding genomes. Public MyIMG annotations on public genomes can be viewed by all users.

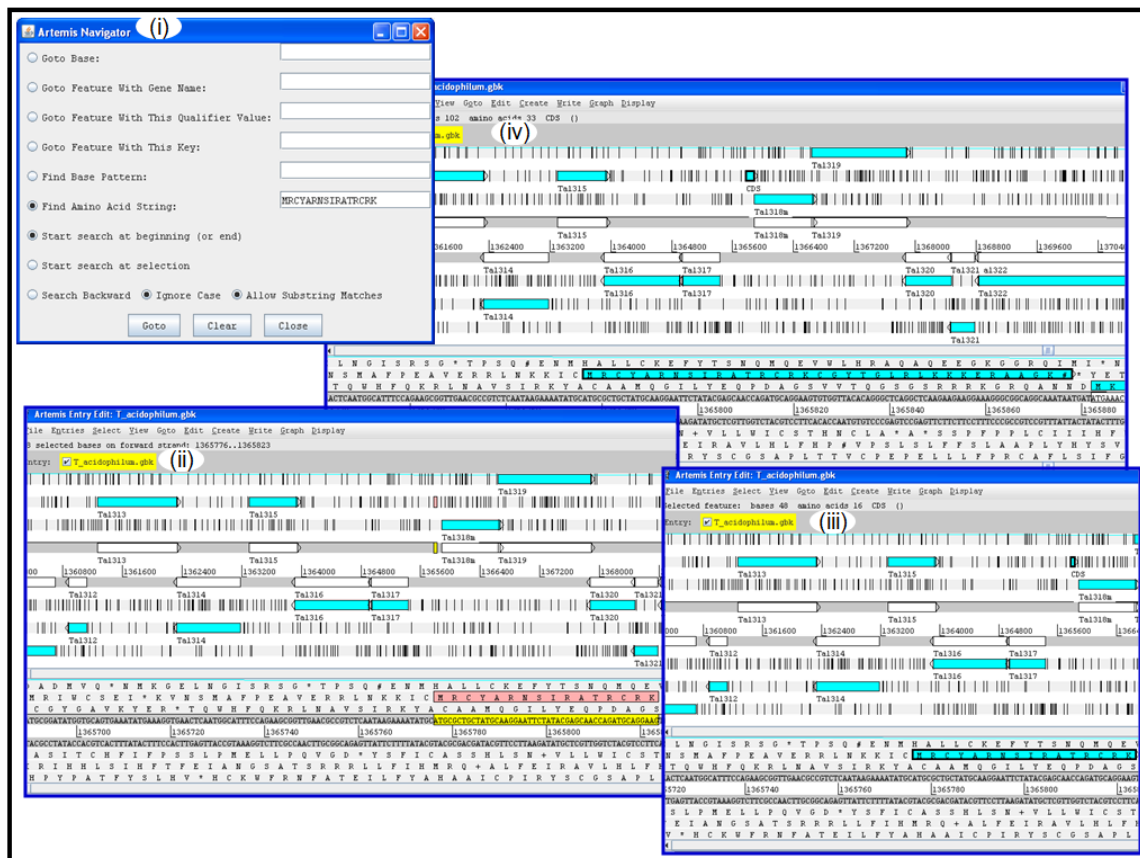
4 Missing Genes

The review of genes and their functional annotations may lead to the identification of missing genes.

4.1 Using Artemis to Find Missing Genes

After determining that a gene x of a genome G is missing because of a similar gene, x' in a closely related genome G' , you can use **Artemis**² to fill in the missing gene as follows:

1. Pick the sequence for gene x' from and run TBLASTn against genome G where you want to find the missing gene.
2. If you get a TBLASTn hit, copy part of the sequence and paste it into the **Artemis** navigator in the box labeled "Find Amino Acid String", as illustrated in Figure 4.1(i). The navigator is under the "Go to" menu. Then click on the "Goto" button.
3. The amino acid sequence is now highlighted, as illustrated in Figure 4.1(ii). Go to the "Create" menu and select "Create feature from base range", as illustrated in Figure 4.1(iii).
4. To extend the gene, go to the "Edit" menu and select "Extend to next stop codon", then select "Fix stop codons".



² Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-A, Barrell B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* 16 (10): 944-945.

Figure 4.1. Using **Artemis** to define a missing gene.

- To find the 5' end, under the "**Edit**" menu, click on "**Extend to previous stop codon**" (you can also use Control-Q for this), as illustrated in Figure 4.2(i).
- To get the amino acid sequence, go to the View menu and select "View amino acid sequence as FASTA". BLAST the sequence against NCBI or IMG. Based on the aligned sequences, find where the start codon should be, as illustrated in Figure 4.2 (ii).
- Select the start codon by pressing "Control-Y". "Control-Y" moves the 5' end to the next potential start codon, as illustrated in Figure 4.2(iii).



Figure 4.2. Using **Artemis** to define a missing gene (cont.).

4.2 Viewing and Adding Missing Genes into IMG

After users identify missing genes in certain genomes, they can add missing genes into IMG. In the **IMG User Annotations** page (Figure 3.1), there is a **View My Missing Gene Annotations** section. Click the "View My Missing Genes" button to view all your missing genes (Figure 4.3(i)).

My Missing Gene Annotations (i)

Select	Taxon Display Name	Count
<input checked="" type="checkbox"/>	Hyperthermus butylicus DSM 5456 [A][F]	2
<input checked="" type="checkbox"/>	Methanococcoides burtonii DSM 6242 [A][F]	1

View Missing Gene Annotations

My Missing Gene Annotations for Selection

Filter column: Genome Name Filter: text

Export Page 1 of 1 << first < prev 1 next > last

Column Selector Select Page Deselect Page

Select	Genome Name	Missing Gene ID	Gene Name
<input type="radio"/>	Hyperthermus butylicus DSM 5456 [A][F]	343	Hyperthermus butylicus DSM 5456 [A][F]
<input type="radio"/>	Hyperthermus butylicus DSM 5456 [A][F]	622	Hyperthermus butylicus DSM 5456 [A][F]
<input type="radio"/>	Methanococcoides burtonii DSM 6242 [A][F]	227	Methanococcoides burtonii DSM 6242 [A][F]

Export Page 1 of 1 << first < prev 1 next > last >> All

[Update Missing Gene Annotation](#)
[Delete Missing Gene Annotation](#)
[Add Missing Gene Annotation](#)
[Add/Update IMG Term\(s\)](#)

My Gene Detail (iii)

Gene Information

Gene ID	343
Gene Symbol	
Locus Type	CDS
Locus Tag	test
Product Name	(1->4)-alpha-D-glucan branching enzyme/(1->4)-alpha-D-glucan synthase (ADP-glucose)
Is Pseudogene?	No
Genome	Hyperthermus butylicus DSM 5456
DNA Coordinates	<2000..2300,3000..>3500 (+)(802bp) -- partial gene
Scaffold Source	Hyperthermus butylicus DSM 5456: NC_008818
Is Public?	No
Protein Information	
Amino Acid Sequence Length	267aa

Figure 4.3. View My Missing Genes

Click "View Missing Gene Annotations" in **My Missing Gene Annotations** page will show a list of existing missing genes (Figure 4.3(ii)). Click on a specific Missing Gene ID will lead to the **My Gene Detail** page of that missing gene (Figure 4.3(iii)).

To add a missing gene, click the "Add Missing Gene Annotations" button to select the target genome, and then click the "Select to Add My Missing Gene" button.

A user can define the following properties of a missing gene:

- Product Name (required)
- Locus Type (required): CDS, tRNA, rRNA, miscRNA, or misc_feature
- Locus Tag (required)
- EC number: for enzyme(s)
- Scaffold
- DNA Coordinates (required):
 - Use comma to separate DNA coordinate ranges; e.g., 3146..3680,5982..8922

20

- Use '<' or '>' to indicate DNA coordinates in partial gene; e.g., <1..30,25..>75
- Strand: + or -
- Is Pseudo Genes? Yes or No
- Description
- Gene Symbol
- Hit Gene ID: if gene has any hits
- Is Public? Yes or No
- Replacing Gene(s): existing genes on the same scaffold to be replaced by this new missing gene

New Missing Gene Annotation

hint: (1) Fields with (*) are required.
(2) Use comma to separate DNA coordinate ranges; e.g., 3146..3680,5982..8922
(3) Use '<' or '>' to indicate DNA coordinates in partial gene; e.g., <1..30,25..>75
(4) Missing genes with 'Is Public?' set to Yes will be visible to all users, *and may be modified by JGI experts.*
(5) If this missing gene is replacing one or more existing genes, enter Gene OID(s) in the 'Replacing Gene(s)' field (comma delimited).

Aquiflexum balticum BA160, DSM 165

Product Name (*)	hypothetical g
Locus Type (*)	CDS
Locus Tag (*)	Aquba_3556
EC Number	
Scaffold	Aquba Con
DNA Coordinates (*)	4139895..414
Strand	+ ▼
Is Pseudo Gene?	No <input type="radio"/> Yes <input checked="" type="radio"/>
Description	
Gene Symbol	
Hit Gene ID	
Is Public?	No <input type="radio"/> Yes <input checked="" type="radio"/>
Replacing Gene(s)	

My Missing Gene Neighborhood

My Gene Id: 0
Start Coord. 4139895
End Coord. 4144045
Strand. +

red = Current Gene
cyan or dashes = My Gene
white = Pseudo Gene
||||| CRISPR array

View Neighborhood

Add My Missing Gene Annotation

Figure 4.4. Adding New Missing Gene Annotation

The "View Neighborhood" button in the **New Missing Gene Annotation** page enables users to graphically view the new missing gene on the selected scaffold in **My Missing Gene Neighborhood** popup (Figure 4.4).

Users can also add IMG terms to the missing genes. To add or update IMG terms of a missing gene, first select the missing gene in the **My Missing Annotations for Selected Genomes** window (Figure 4.3 (ii)) and then click the "Add/Update IMG Term(s)" button. Select proper IMG term(s), and then click the "Update Term Association" button to update.

4.3 Group and Community Missing Gene Annotations in IMG

Similar to MyIMG annotations, missing gene information can be shared among group members. In order for a user to share missing gene information with group members, the user should first click the "View My Missing Genes" button in Figure 3.1. The new "Update Missing Gene Sharing in Selected Genome(s)" section allows updates to group sharing options (Figure 4.5).

My Missing Gene Annotations

Select	Taxon Display Name	Count
<input checked="" type="checkbox"/>	Hyperthermus butylicus DSM 5456 [A][F]	2
<input checked="" type="checkbox"/>	Methanococcoides burtonii DSM 6242 [A][F]	1

View Missing Gene AnnotationsAdd Missing Gene AnnotationSelect AllClear All

Update Missing Gene Sharing in Selected Genome(s)

Share or remove sharing of all missing genes in selected genome(s). (Note: You will have to select all groups included in your previous share be removed.)

Option: ☒ Remove group sharing for all my missing gene annotations
☐ Share all missing genes with selected group(s)
☐ GBP
☐ ProPortal

Update Sharing

Figure 4.5. Sharing Missing Gene Annotations of Selected Genomes

To share individual missing genes, a user should click the "View Missing Gene Annotations" button (in Figure 4.5) to view the individual missing gene list. A new "Update Group Sharing" section allows users to change the group sharing option of selected gene.

Missing gene annotations are private by default. A user can change his/her missing gene annotations to public. Public missing gene annotations are visible to all users who have access permission to the corresponding genomes. Moreover, since IMG ER is an "Expert Review" site, all public missing gene annotations can be reviewed and modified by JGI experts.

